

## Appendix I

```

>gb|AAL78656.1|AF472514_1 G thymus high mobility group box protein TOX [Mus
musculus]
Length=526

GENE ID: 252838 Tox | thymocyte selection-associated HMG box gene
[Mus musculus] (Over 10 PubMed links)

Score = 293 bits (749), Expect = 2e-78, Method: Compositional matrix adjust.
Identities = 183/305 (60%), Positives = 226/305 (74%), Gaps = 16/305 (5%)

Query 1 MSDGNPELLSTTSQYNSQGESNEDYEIPPIPPNLPEPSLLHLGDEAGYHSLCHGLAPN 60
M++ + + + SQ+Y +ED+ IPPITPP+LP+ SL+HL + E+GYHSLCH + N
Sbjct 44 MTEPSQDYVPASQSYFGPSLESEDFNIPPITPPSLPDHSLVHLNEVESGYHSLCHPMNHN 103

Query 61 GLLPAYSYQAMDLPAIMVSNMLAQDGHLLSGQLPTIQEMVHSEVAAYDS-----G 110
GLLP + Q MDLP I VSNML QDG LLS + +QE+ ++E A Y S G
Sbjct 104 GLLP-FHPQIMDLPEITVSNMLGQDQALLSNSISVMQEIGNAEGAQYSSHPQMAAMRPRG 162

Query 111 RPGPLLRPAMLA-SHMGALSQSQSLISQMGLRSG---IAHSSPPPGSKSATPSPSSSTQ 166
+P + + +M+ ++ ++QSQL +Q+GL G +AH+SPSPPGSKSATPSPSSS
Sbjct 163 QPTDIRQQASMMQPGQLTTINQSQLSAQLGLNMGGINVAHNSPSPPGSKSATPSPSSSVH 222

Query 167 EEESDAHFKISG-EKRPSTDGPKKAKNPKKKKKKDPNEPQKPVSAAYALFFRDTQAAIKGQ 225
E+E + KI+G EKRP++D GKK K PKKKKKKDPNEPQKPVSAAYALFFRDTQAAIKGQ
Sbjct 223 EDECEDASKINGGKRPASDMGKKPKTPKKKKKKKDPNEPQKPVSAAYALFFRDTQAAIKGQ 282

Query 226 NPATFGDVSKIVASMWDSLGEQKQAYKRKTEAAKKEYLKALAAAYRASLVSKSPDPQGE 285
NP+ATFG+VSKIVASMWDLGEEQKQ YK+KTEAAKKEYLK LAAYRASLVSKS D +
Sbjct 283 NPATFGEVSKIVASMWDLGEEQKQVYKKKTEAAKKEYLKQALAAAYRASLVSKSYTDPVD 342

Query 286 AKNAQ 290
K +Q
Sbjct 343 VKTSQ 347

```

Query: SEQ ID NO: 1